

Manpower forecasting and the aggregation problem: a microeconomic approach

Citation for published version (APA):

van Eijs, P. W. L. J., & Borghans, L. (1993). *Manpower forecasting and the aggregation problem: a microeconomic approach*. Researchcentrum voor Onderwijs en Arbeidsmarkt, Faculteit der Economische Wetenschappen. ROA Research Memoranda No. 6E <https://doi.org/10.26481/umaror.199306E>

Document status and date:

Published: 01/01/1993

DOI:

[10.26481/umaror.199306E](https://doi.org/10.26481/umaror.199306E)

Document Version:

Publisher's PDF, also known as Version of record

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

MANPOWER FORECASTING AND THE AGGREGATION
PROBLEM: A MICROECONOMIC APPROACH

ROA-RM-1993/6E

Patrick van Eijs and Lex Borghans

RESEARCH CENTRE FOR EDUCATION AND THE LABOUR MARKET

Faculty of Economics and Business Administration
Rijksuniversiteit Limburg

Maastricht, November 1993

CIP-GEGEVENS KONINKLIJKE BIBLIOTHEEK, DEN HAAG

Eijs, Patrick van

Manpower forecasting and the aggregation problem: a microeconomic approach/Patrick van Eijs and Lex Borghans. - Maastricht: Research Centre for Education and the Labour Market, Faculty of Economics and Business Administration, Rijksuniversiteit Limburg. - ([Report] / Research Centre for Education and the Labour Market, ISSN 0922-8098; ROA-RM-1993/6E)

Met lit. opg.

ISBN 90-5321-112-8

Trefw.: arbeidsmarkt

CONTENTS

	Page
ABSTRACT	i
1. INTRODUCTION	1
2. THE RAS MODEL	4
3. AN ECONOMIC FOUNDATION OF RAS	6
3.1. Introduction	6
3.2. Education as an input-factor	6
3.3. The allocation model	7
4. TWO HYPOTHESES	11
5. RAS VARIANTS	16
6. SOME PROBLEMS WITH RAS FORECASTING	20
6.1. The multiple solution problem	20
6.2. The empty cells problem	20
7. EMPIRICAL ANALYSIS	24
7.1. Measurement of the performance	24
7.2. Analysis of the performances	25
7.3. The competitiveness hypothesis: further indications	26
7.4. The aggregation problem	30
8. CONCLUSIONS	31
REFERENCES	33
APPENDIX 1. INDUSTRIAL AGGREGATION LEVELS	34
APPENDIX 2. PROOF OF PROPOSITION 1	35
APPENDIX 3. PROOF OF PROPOSITION 2	36

ABSTRACT

This paper deals with the aggregation issue in manpower forecasting. In particular, it deals with investigating the opportunities to aggregate data over industries in an occupation by education framework. Starting point is a microeconomic allocation model of the firm in which the optimal employment by education is determined. Two hypotheses dealing with the uniformity of wage changes and technologies over industries are formulated. Several variants of the allocation model differing with regard to accepting none, one or both hypotheses are investigated. It is shown that these variants can all be re-arranged to obtain the so-called RAS-structure.

The performances of the RAS variants give an indication of the validity of the hypotheses. Two factors determine the performance (1) the validity of the hypotheses and (2) the quality of the data. It is concluded that in the Dutch situation there is no empirical justification for incorporating wage-differentials between different industries for one type of education. In these data there is no need to distinguish differences in the production function between the industries also. This empirical observation is much less strong, however.

We would like to thank Ron Dekker, Andries de Grip, Hans Heijke and Ed Willems for their helpful comments.

1. INTRODUCTION

Manpower forecasting deals with projecting the labour demand and supply, especially the occupational and educational structure of demand and supply. Its purpose is to give insight into the future developments at the labour market, which enables participants to make necessary adjustments to the changing situation in time. To make manpower forecasts suited for this kind of decision making, it is very important to distinguish different groups at the labour market, if these groups might be distinguished with regard to labour market developments. Therefore, the level of aggregation is an important problem in manpower forecasting.

During the last forty years a lot of manpower forecasting techniques have been developed. According to Corcoran, Hughes and Sexton (1993) a distinction can be made between single-cell and matrix models. Given the matrix describing the current labour market structure, single-cell models investigate developments in every cell of the matrix separately. The best-known single-cell technique is the manpower requirements approach.¹ Matrix models differ from single-cell models in taking spill-over effects between cells into account. The most important representative of the matrix models is the so-called *RAS model*. It has been shown that RAS yields rather good forecasts compared to other forecasting techniques.²

At first sight the RAS model seems to be a purely mechanical approach. However, several attempts have been made to give an economic foundation of the RAS model. One of these is made by Evans and Wabe (1974), who apply RAS to an industry by occupation matrix. They have used a microeconomic allocation model to show that, departing from a Cobb-Douglas production technology and perfect competition, RAS generates vectors which can be interpreted as changes in wages and product prices. This finding is the starting point for this paper.

In this paper it is shown that an adjusted variant of the allocation model given by Evans and Wabe offers the opportunity to investigate the functioning of the labour market at a low level of aggregation. In particular, by dividing the economy into different industries, attention is paid to two hypotheses concerning the aggregation problem.

The first hypothesis concerns the competitiveness of the labour market. The allocation model of Evans and Wabe assume the wage changes to be uniform over the market. This implies that wage changes are not industry-specific. This first hypothesis is called the *competitiveness hypothesis*. The second hypothesis concerns the production technology of the firms. In this paper, the production technology is described by a two-stage production function. The first stage describes how the industrial production is generated by capital and occupational activities. The second stage describes the share of the types of education in the occupational

1. For an overview of the models in the manpower requirements tradition, see Van Eijs (1993).

2. See, for example, Evans and Lindley (1973), Bosworth, Evans and Lindley (1974) and Borghans and Heijke (1993).

activities. The *technology hypothesis* of the allocation model assumes that both stages of the production technology are industry-specific. In this paper this technology hypothesis is tested by investigating the industry-specificity of the production function of occupational activities (that is, the second stage of the production function).

To investigate the aggregation problem an allocation model is presented in which both hypotheses are dropped. Wages are not uniform over industries and the occupational production technology is industry-specific. Furthermore, in this standard model in all industries every occupational activity has one specific production function and wage changes are uniform for every type of education over occupational classes within an industry. It will be shown that the changes in the wages in this standard model can be calculated by applying the RAS method to the matrix occupation by education for each industry separately.

The competitiveness and the technology hypothesis lead to two variants of the standard model. In the first variant only the competitiveness hypothesis is accepted. In this variant the wage changes for each type of education are assumed to be uniform over all industries, reflecting the competitiveness of the labour market. However, the production technology is still assumed to be industry-specific. In the second variant both hypotheses are accepted. It is assumed that both wage changes are uniform and the occupational production technology is similar in each industry for every occupational activity. It is shown that the first and the second variant both can be represented by an adjusted formulation of the RAS procedure. In the first variant the RAS procedure is applied to the matrix occupation in a particular industry by education, while in the second variant the procedure is applied to the aggregated matrix occupation by education. Furthermore, these variants are applied to five aggregation levels. The two extremes are (1) considering the economy to be one industry and (2) dividing up the economy into 14 industries. The standard RAS model can be considered to be the special case of the second variant in which the lowest aggregation level is applied. In a third possible variant only the technology hypothesis could be accepted. This variant is not investigated because it does not offer the opportunity to identify differences in production technologies by means of RAS.

The problem discussed in this paper has both theoretical and practical relevance. Firstly, from a theoretical point of view, the validity of the competitiveness hypothesis and the technology hypothesis is investigated. The analysis therefore gives some evidence on the competitiveness of the labour market. Possibly, conclusions on the competitiveness can be drawn with regard to educational submarkets. Furthermore, it gives some evidence on the extent to which production processes are industry-specific. Secondly, from a practical point of view, the analysis can be a guideline for decisions on the optimal aggregation level of future manpower forecasts by means of the RAS model. The empirical analysis shows which aggregation level yields the best results.

The structure of the remainder of this paper is as follows. Section 2 pays attention to the RAS model. Section 3 describes the standard allocation model which is a variant of Evans's and Wabe's model. It is shown that this standard model can give an economic interpretation of RAS. In section 4 the two hypotheses mentioned above are worked out. It is shown that, if

these hypotheses are accepted, the allocation model still can be re-arranged to obtain RAS. Section 5 describes the two RAS variants mentioned above and pays attention to the links between these variants and the hypotheses. Section 6 pays attention to two technical problems associated with RAS forecasting. The first problem deals with the fact that RAS has multiple solutions. The second problem deals with the fact that RAS sometimes does not have a solution at all. Section 7 discusses the results of the empirical analysis for the Dutch economy. Finally, in section 8 some concluding remarks are made.

2. THE RAS MODEL

Much attention has been paid to instruments that can contribute to the avoidance of labour market mismatches. One of these instruments is manpower forecasting. Manpower forecasting deals with projecting the labour demand and supply, especially the occupational and educational structure of demand and supply. The RAS model has been introduced in manpower forecasting literature in the 1960s.³ Starting point of the RAS model is a matrix describing the current labour market situation. In manpower forecasting applications this often is an industry by occupation matrix. Other possibilities are an industry by education or an occupation by education matrix.

In this paper RAS is applied to an occupation by education framework.⁴ This implies that the aim of the RAS model is to predict a future occupation by education matrix, based on the present distribution of types of education over occupations. The model uses forecasts of the row and column totals and aims to bring the current matrix in alignment with these future row and column totals.

The model looks as follows. RAS searches for vectors r and s such that:

$$c_{ij} = r_i a_{ij} s_j \quad \forall i, j \quad (1)$$

satisfies the following constraints:

$$\sum_j c_{ij} = b_i^{occ} \quad \forall i \quad (2)$$

$$\sum_i c_{ij} = b_j^{educ} \quad \forall j \quad (3)$$

in which a_{ij} and c_{ij} are elements of the base year matrix A and the forecasted matrix C (which is the forecast of the target year matrix B) respectively, b_i^{occ} is the i th row total of the future matrix B and b_j^{educ} is the j th column total of B . From a manpower forecasting point of view this implies that for a particular year the forecasts for total employment per type of education are given and that these forecasts, starting from the distribution in a base year, have to be translated into the distribution of workers with particular qualifications (type of education) over the occupations.

3. See, for example, Brown, Leicester and Pyatt (1964).

4. See, for an explanation of this choice, section 3.

It has been shown that the solution of (1) - (3) can be attained by an iterative process. This RAS procedure consists of the following steps. Firstly, the current matrix is adjusted for the future row totals, to obtain the first cell-forecasts c_{ij}^1 :⁵

$$c_{ij}^1 = \frac{b_i^{occ}}{a_i^{occ}} \cdot a_{ij} \quad \forall j \quad (4)$$

in which a_i^{occ} is the i th row total of the current matrix.

After this first step, the iterative process is started during which the column total constraints (3) and the row total constraints (2) are violated in turn. First, the column constraints are satisfied (and the row constraints are violated) by post-multiplying the matrix by the ratio between the future column totals b_j^{educ} and the column totals of the estimated matrix $c_j^{educ,1}$, to obtain the second cell-forecasts c_{ij}^2 :

$$c_{ij}^2 = c_{ij}^1 \cdot \frac{b_j^{educ}}{c_j^{educ,1}} = \frac{b_i^{occ}}{a_i^{occ}} \cdot a_{ij} \cdot \frac{b_j^{educ}}{c_j^{educ,1}} \quad \forall i \quad (5)$$

Secondly, the roles are reversed; the row constraints are satisfied by pre-multiplying the new forecasts c_{ij}^2 by the ratio between the future row totals b_i^{occ} and the row totals of the estimated matrix $c_i^{occ,2}$:

$$c_{ij}^3 = \frac{b_i^{occ}}{c_i^{occ,2}} \cdot c_{ij}^2 = \frac{b_i^{occ}}{c_i^{occ,2}} \cdot \frac{b_i^{occ}}{a_i^{occ}} \cdot a_{ij} \cdot \frac{b_j^{educ}}{c_j^{educ,1}} \quad \forall j \quad (6)$$

These last two steps are repeated in an iterative process until the forecasted matrix converges and satisfies both sets of constraints at the same time. If a solution for (1) - (3) exists, this iterative procedure always converges to this solution, with $(c_i^{occ,0} = a_i^{occ})$:

$$r_i = \prod_{n=0}^{\infty} \frac{b_i^{occ}}{c_i^{occ,2n}} \quad \forall i \quad (7)$$

and:

$$s_j = \prod_{n=1}^{\infty} \frac{b_j^{educ}}{c_j^{educ,2n-1}} \quad \forall j \quad (8)$$

5. The other option is to start with the adjustment for column totals. This will yield the same cell-forecasts.

3. AN ECONOMIC FOUNDATION OF RAS

3.1. Introduction

Several options for an economic foundation of the RAS model have been suggested. For example, Evans and Lindley (1973) mention (and refute) an economic explanation of the fact that the RAS model minimises the 'distance' between the matrices A and C , which is the forecast of B . They pay attention to the suggestion that this feature of RAS implies that the labour market tries to minimise the shifts in labour force composition.⁶ Cohen (1988) also tries to give an economic rationale of the RAS model in an occupation by education framework. He suggests that the demand side of the labour market determines the occupational structure whereas the supply side determines the educational structure. This implies that imbalances are likely to occur. The iterative process of RAS reflects the trail and error process at the labour market in order to obtain a new labour market equilibrium. Finally, Borghans and Heijke (1993) use a model similar to RAS to introduce substitution effects into the manpower requirements approach.

The multiplying vectors r and s are sometimes interpreted as measures for the size of the substitution effects. Evans and Wabe (1974) go even further. Departing from a Cobb-Douglas production technology and perfect competition, they proof that, in an industry by occupation framework, the vectors r and s can be considered to be vectors of, respectively, wage and product price changes. In this paper, their analysis is used to investigate the validity of two hypotheses by means of the RAS model. These hypotheses deal with the competitiveness of the labour market and the industry-specificity of production processes. The remainder of this section describes the standard allocation model. In section 4 the two hypotheses are added to this standard model.

3.2. Education as an input-factor

Evans and Wabe (1974) use an industry by education matrix to proof that the vectors r and s can be considered to be vectors of product price and wage changes. Wages are in their model distinguished at occupational level. As already mentioned in section 2, in this paper wages are treated as characteristics of types of education, which is done for the following reasons. In the first place, since the educational background of people rather than the occupation they have seems to be an input-factor in the production process, from a competition point of view, it seems to be reasonable to assume that wages are linked with types of education instead of occupations. This implies that allocation processes at the labour market take place by way of types of education. Secondly, from a manpower forecasting point of view, it is important to distinguish homogenous groups in an allocation model at a level at which crucial choices are to

6. Although the RAS model does not minimise the number of persons that move from one job to another, but minimises the entropy distance measure $\sum x_{ij} \log x_{ij}$ with $x_{ij} = c_{ij} - a_{ij}$.

be made. In this respect educational investment decisions are much more crucial than the occupational choice. It seems to be reasonable to consider a type of education, which is an input, to a homogenous group for which a uniform price (the wage) has to be paid.

3.3. The allocation model

Suppose every industry consists of one representative profit-maximising firm.⁷ This implies that, in an industry k ($k = 1..o$), the production technology can be described by the following industry-specific Cobb-Douglas production function:

$$Y_k = v_k (CP_k)^{\alpha_k} \prod_i (y_{ik})^{\beta_{ik}} \quad (9)$$

There are $m+1$ 'production factors': capital (CP) and m occupational activities i . Equation (9) shows that these activities contribute y_{ik} to the total industrial production Y_k . y_{ik} is produced by workers who can be characterised by their qualifications (type of education j). The Cobb-Douglas 'occupational production function' describing the occupational activities looks as follows:

$$y_{ik} = \mu_{ik} \prod_j (L_{ijk})^{\gamma_{ijk}} \quad (10)$$

in which L_{ijk} is the number of workers having qualifications j ($j = 1..n$), working in occupational activity i in industry k . Note that also this occupational activity production function is industry-specific, which implies that in each industry, occupations have a particular contents, requiring workers having specific qualifications. Furthermore, this production functions implies that the productivity of a worker not only depends on his qualifications, but also on both the occupational activity and the industry in which he or she is employed.

Substituting (10) into (9) yields a model comparable to Evans's and Wabe's model. However, there is one major distinction: Evans's and Wabe's model refers to an industry by occupation matrix, while this model then refers to an industry by education matrix, because in contradiction with Evans's and Wabe's model, wages are linked with types of education instead of occupations.

The distinction between two production functions seems to be artificial. However, it can be shown that this specification of the production process implies that the underlying decision making process consists of two independent stages. Firstly, the optimal allocation of resources over capital and the occupational activities is determined, given the price of capital and the shadow prices of the occupational activities. Secondly, the optimal allocation of resources within an occupational activity over the various types of education is determined. This implies

7. This assumption is not very restrictive. Later it will be shown that if firms have the same technology, aggregation of these firms results in a similar Cobb-Douglas production technology.

that there are two possibilities to formulate the profit function. Firstly, an overall profit function and secondly a two-stage profit function, in which every occupational activity maximises his own profit. Both possibilities yield the same results.⁸ The one-stage formulation is used here. This implies that the profit maximising problem looks as follows:

$$\begin{aligned} \max \pi_k &= P_k Y_k - \sum_j \sum_i w_{ijk} L_{ijk} - R_k C P_k \\ \text{s.t. } Y_k &= v_k (C P_k)^{\alpha_k} \prod_i (y_{ik})^{\beta_{ik}} \\ y_{ik} &= \mu_{ik} \prod_j (L_{ijk})^{\gamma_{ijk}} \end{aligned} \quad (11)$$

in which π_k is a firm's profit in industry k , P_k is the price per unit of output in industry k and R_k is the price of capital. Furthermore, it is assumed that the wage w_{ijk} is paid to workers according to their qualifications (type of education).⁹ This profit maximisation problem can be solved by means of the Lagrange method. The Lagrangian La looks as follows:

$$\begin{aligned} La &= P_k Y_k - \sum_j \sum_i w_{ijk} L_{ijk} - R_k C P_k - \\ &\quad \lambda_k \left[Y_k - v_k (C P_k)^{\alpha_k} \prod_i (y_{ik})^{\beta_{ik}} \right] - \\ &\quad \lambda_{ik} \left[y_{ik} - \mu_{ik} \prod_j (L_{ijk})^{\gamma_{ijk}} \right] \end{aligned}$$

The first order conditions of this problem are:

$$\frac{\partial La}{\partial Y_k} = P_k - \lambda_k = 0 \quad (12a)$$

$$\frac{\partial La}{\partial C P_k} = -R_k + \lambda_k \alpha_k \frac{Y_k}{C P_k} = 0 \quad (12b)$$

$$\frac{\partial La}{\partial y_{ik}} = \lambda_k \beta_{ik} \frac{Y_k}{y_{ik}} - \lambda_{ik} = 0 \quad (12c)$$

$$\frac{\partial La}{\partial L_{ijk}} = -w_{ijk} + \lambda_{ik} \gamma_{ijk} \frac{y_{ik}}{L_{ijk}} = 0 \quad (12d)$$

$$\frac{\partial La}{\partial \lambda_k} = Y_k - v_k (C P_k)^{\alpha_k} \prod_i (y_{ik})^{\beta_{ik}} = 0 \quad (12e)$$

$$\frac{\partial La}{\partial \lambda_{ik}} = y_{ik} - \mu_{ik} \prod_j (L_{ijk})^{\gamma_{ijk}} = 0 \quad (12f)$$

8. In demand theory, this feature is known as separability.

9. It is assumed here that workers are paid according to their qualifications obtained by formal education.

As mentioned above this decision process can be separated into two stages. In this paper the focus is on the second stage to which the equations (12d) and (12f) refer. Suppose that the wage paid to a worker having qualifications j is uniform over all occupational activities within an industry:

$$w_{ijk} = w_{jk} \quad \forall i \quad (13)$$

If this is not the case, workers move from one occupational activity to another. Substituting (13) into (12d) and re-arranging yields:

$$L_{ijk} = \frac{\gamma_{ijk} \lambda_{ik} y_{ik}}{w_{jk}} \quad (14)$$

in which the Lagrangian multiplier λ_{ik} can be interpreted as the shadow price of a unit of occupational output y_{ik} . This is the labour demand function for workers having qualifications j in occupational activity i and industry k . Suppose the occupational production technology does not change over time.¹⁰ This implies that γ_{ijk} is constant over time. By introducing a time dimension and writing down equation (14) for two points in time (t and $t-1$), the following equation can be derived:

$$\frac{L_{ijk}^t}{L_{ijk}^{t-1}} = \left(\frac{y_{ik}^t}{y_{ik}^{t-1}} \right) \left(\frac{\lambda_{ik}^t}{\lambda_{ik}^{t-1}} \right) \left(\frac{w_{jk}^{t-1}}{w_{jk}^t} \right) \quad (15)$$

To put it in words: the relative change in employment in industry k in occupational activity i for workers having qualifications j is the product of the relative change in production by occupational activity i , the relative change in the shadow price of occupational activity i and the (inverse) relative wage change. Re-arranging (15) yields:

$$L_{ijk}^t = \left(\frac{y_{ik}^t}{y_{ik}^{t-1}} \right) \left(\frac{\lambda_{ik}^t}{\lambda_{ik}^{t-1}} \right) L_{ijk}^{t-1} \left(\frac{w_{jk}^{t-1}}{w_{jk}^t} \right) \quad (16)$$

Finally, combining (1) and (16) and applying RAS to the industry by occupation by type of education matrix L yields:

$$r_{ik} = \left(\frac{\lambda_{ik}^t}{\lambda_{ik}^{t-1}} \right) \left(\frac{y_{ik}^t}{y_{ik}^{t-1}} \right) \quad (17)$$

10. Evans and Wabe do not make this assumption explicitly, although it is essential to derive the most important equation in the model (equation (15) in this paper).

and:

$$s_{jk} = \left(\frac{w_{jk}^{t-1}}{w_{jk}^t} \right) \quad (18)$$

So, according to this allocation model, the vector r can be interpreted as a vector of the product of the change in the shadow price and the change in the contribution of occupational activity i to the total industrial production. If $r_{ik} > 1$, the value contribution of occupational activity i in industry k has risen. If $r_{ik} < 1$, the value contribution of occupational activity i in industry k has fallen. The vector s can be interpreted as a vector of wage changes. If $s_{jk} > 1$, the wage paid for qualifications j has fallen. If $s_{jk} < 1$, the wage paid for qualifications j has risen. It has to be noted here that the term wage in this context should be taken broadly as an indication of market scarcity.

This is the standard model. Note that both the vector r and the vector s are industry-specific. This implies that both vectors can be calculated by applying RAS to each industry separately. In the next section, two restricting assumptions are imposed on this standard allocation model. It is shown how the RAS model can be used to investigate the validity of these two assumptions.

4. TWO HYPOTHESES

The standard allocation model leads to a RAS model in which the multiplying factors are industry-specific. The following restrictive assumptions can be imposed:

Assumption 1

$$w_{ijk} = w_j \quad \forall i, k \quad (19)$$

Assumption 2

$$\mu_{ik} = \mu_i \quad \forall k \quad (20a)$$

$$\gamma_{ijk} = \gamma_{ij} \quad \forall k \quad (20b)$$

Assumption 1 states that wages changes are uniform over the whole economy for every type of education. This implies that wage changes are uniform also. In fact, assumption 1 consists of two parts: (1) wage changes are uniform over all occupational classes within an industry and (2) wage changes are uniform over all industries. The first part is indissolubly linked with RAS. The vector s implies that the multiplying factor s_j (that is, the wage change for type of education j) is equal for all elements in the corresponding column of the matrix (that is, for all occupational classes). Therefore, this part of the first assumption is already imposed in the standard model (equation (13)). If the second part of the assumption is imposed on the allocation model, equation (18) looks as follows:

$$s_j = \left(\frac{w_j^{t-1}}{w_j^t} \right) \quad (21)$$

So, to put it in words, the following *competitiveness hypothesis* can be set:

Hypothesis 1:

The labour market is perfect; wage changes are not industry specific.

If the labour market is perfect and in equilibrium, wage changes have to be uniform over all industries (equation (19)). This implies, in the RAS framework, that the vector s is uniform over all industries. This is reflected by equation (21). If, on the contrary, the labour market is imperfect, wage differentials persist and the vector s is not uniform over all industries. This is reflected by equation (18). It is meaningful then to use disaggregated data with regard to types of education in a RAS framework. Column totals do not refer then to aggregated types of education, but to types of education in a particular industry. It is useful to distinguish industrial types of education, because, if aggregated data are used, wage differentials between industries are not taken into account. For example, a distinction is made between people having higher vocational engineering education in, for example, the public services and the chemical industry.

Assumption 2 states that the occupational production technology is not industry specific. This implies that it is assumed that occupations have the same contents in every industry and that the productivity of a worker depends on his qualifications and on the occupational activity in which he or she is employed, but not on the industry in which he or she is employed. Note that, in contradiction to the effect of imposing assumption 1 on the vector s , equation (17) does not change. This implies that the vector r remains industry specific, if assumption 2 is imposed. The following *technology hypothesis* can be set:

Hypothesis 2:

The occupational production technology is not industry-specific.

Suppose the competitiveness hypothesis has already been accepted. As a result types of educations are not treated to be industry-specific. In other words, the column totals are aggregated. In the remainder of this section it is proven that, if the technology hypothesis is accepted subsequently to the competitiveness hypothesis, it is allowed to aggregate the row totals over industries, although the vector r is industry specific. If the technology hypothesis is accepted, the optimal number of workers in occupational activity i having qualifications j is:

$$L_{ijk} = \frac{\gamma_{ij} \lambda_{ik} y_{ik}}{w_j} \quad (22)$$

Aggregating the data over industries, in other words, treating occupational classes to be not industry-specific and accepting the technology hypothesis, yields:

$$L_{ij} = \sum_k L_{ijk} = \sum_k \frac{\gamma_{ij} \lambda_{ik} y_{ik}}{w_j} = \frac{\gamma_{ij}}{w_j} \sum_k \lambda_{ik} y_{ik} \quad (23)$$

Furthermore, the production of occupational activity i y_{ik} is endogenous in this model.¹¹ Substituting (22) into the occupational production function (10) and accepting the technology hypothesis yields:

$$\begin{aligned} y_{ik} &= \mu_i \prod_j \left(\frac{\gamma_{ij} \lambda_{ik} y_{ik}}{w_j} \right)^{\gamma_{ij}} \rightarrow \\ y_{ik} &= \mu_i \lambda_{ik}^{\sum_j \gamma_{ij}} y_{ik}^{\sum_j \gamma_{ij}} \prod_j \left(\frac{\gamma_{ij}}{w_j} \right)^{\gamma_{ij}} \rightarrow \\ y_{ik} &= \mu_i^{\frac{1}{1-\sum_j \gamma_{ij}}} \lambda_{ik}^{\frac{\sum_j \gamma_{ij}}{1-\sum_j \gamma_{ij}}} \prod_j \left(\frac{\gamma_{ij}}{w_j} \right)^{\frac{\gamma_{ij}}{1-\sum_j \gamma_{ij}}} \end{aligned} \quad (24)$$

11. Evans and Wabe do not mention this characteristic of the model.

Let:

$$\theta_i = \frac{1}{1 - \sum_j \gamma_{ij}} \text{ with } \sum_j \gamma_{ij} \neq 1$$

It is important to note that, only if the technology hypothesis is accepted, θ_i is not industry-specific. Furthermore, θ_i does not depend on the type of education. Substituting into (24) yields:

$$y_{ik} = \mu_i^{\theta_i} \lambda_{ik}^{\theta_i-1} \prod_j \left(\frac{\gamma_{ij}}{w_j} \right)^{\gamma_{ij} \theta_i}$$

Let further:

$$W_i^{ind} = \frac{1}{\prod_j \left(\frac{\gamma_{ij}}{w_j} \right)^{\gamma_{ij} \theta_i}} = \prod_j \left(\frac{w_j}{\gamma_{ij}} \right)^{\gamma_{ij} \theta_i}$$

W_i^{ind} is a weighted wage index. The more 'important' the type of education j in the occupational production y_{ik} (that is, the higher γ_{ij}), the higher the weight factor of w_j . W_i^{ind} can be interpreted as a cost price index of the production by occupational activity i . Here the link between the commodity production function (9) and the occupational production function (10) becomes clear. The higher the cost price index W_i^{ind} of occupational activity i , ceteris paribus, the more occupation i is substituted for the other occupational activities and capital in the production of commodity k . It is important to note that, only if the technology hypothesis is accepted, W_i^{ind} is not industry-specific. Now equation (24) can be written as:

$$y_{ik} = \frac{\mu_i^{\theta_i} \lambda_{ik}^{\theta_i-1}}{W_i^{ind}} \quad (25)$$

which can be interpreted as the demand equation for occupational activity i in industry k . Substituting (25) into (23) yields:

$$\begin{aligned} L_{ij} &= \frac{\gamma_{ij}}{w_j} \sum_k \frac{\mu_i^{\theta_i} \lambda_{ik}^{\theta_i}}{W_i^{ind}} \rightarrow \\ L_{ij} &= \gamma_{ij} \mu_i^{\theta_i} \left(\frac{1}{w_j} \right) \left(\frac{1}{W_i^{ind}} \right) \sum_k \lambda_{ik}^{\theta_i} \end{aligned} \quad (26)$$

Equation (26) is an aggregated equivalent of demand equation (14).¹² This implies that, assuming again that the production technology does not change over time, equation (27) can be derived from equation (26) in the same way as equation (16) is derived from equation (14):

12. Of course, in equation (14) y_{ik} can also be replaced by the exogeneous parameters and the shadow price.

$$L_{ij}^t = \left(\frac{W_i^{ind,t-1}}{W_i^{ind,t}} \right) \left(\frac{\sum_k \lambda_{ik}^{\theta_{it}}}{\sum_k \lambda_{ik}^{\theta_{it-1}}} \right) L_{ij}^{t-1} \left(\frac{w_j^{t-1}}{w_j^t} \right) \quad (27)$$

This implies that:

$$r_i = \left(\frac{W_i^{ind,t-1}}{W_i^{ind,t}} \right) \left(\frac{\sum_k \lambda_{ik}^{\theta_{it}}}{\sum_k \lambda_{ik}^{\theta_{it-1}}} \right) \quad (28)$$

and:

$$s_j = \left(\frac{w_j^{t-1}}{w_j^t} \right) \quad (29)$$

Equation (27) shows that, if the competitiveness hypothesis is accepted and if the column totals (that is, total number of workers having the particular qualifications) are aggregated over the industries, applying RAS to, on the one hand, a matrix without and, on the other hand, a matrix with aggregated row totals yield the same vector s . This implies that both variants yield the same forecasts and that the vector s is still interpretable as a vector of wage changes.

Equation (28) shows that the vector r is not industry-specific. It only depends on the occupational class i . The vector r is the product of vectors of wage index changes and of changes in an index of shadow prices. Because r_i can be written as the change in the ratio between the shadow price and the cost price index, it might be interpreted as the change in a kind of value added index of occupational activity i . To put it briefly, if both the competitiveness and the technology hypothesis are accepted, treating both occupational classes and types of education to be not industry-specific, yields the same forecasts as applying RAS to disaggregated data with regard to occupational classes, in which, for example, a distinction is made then between lower administrative occupations in the public services and the chemical industry. In other words, if both hypotheses are accepted, both row and column totals can be aggregated over industries to obtain an aggregated RAS variant, which yields the same forecasts as disaggregated RAS variants.

To summarise this section, if the competitiveness hypothesis is accepted, the vector s is equal for all industries and disaggregating data with regard to types of education and distinguishing 'industrial types of education' makes no sense with regard to making manpower forecasts by means of RAS. If the technology hypothesis is also accepted, disaggregating data with regard to occupational classes and distinguishing 'industrial occupational classes' also makes no sense

with regard to making manpower forecasts by means of RAS.¹³ In the remainder of this paper, these hypotheses are investigated by means of the RAS framework. The next section describes the RAS variants which are used to do this.

13. In this paper a specific order for investigating the hypotheses is chosen. Firstly, the competitiveness hypothesis is considered and after that the technology hypothesis is investigated. If wages for the same type of education differ between industries, it is not possible, by means of the RAS framework, to identify differences in production technologies.

5. RAS VARIANTS

The preceding section made clear that three aspects of the data are relevant with regard to investigating the competitiveness and the technology hypothesis:

1. The degree of industrial disaggregation in general. That is, the number of industries that is distinguished.

Given a decision with regard to this degree of industrial disaggregation, two other decisions have to be made:

2. Should a distinction be made between the same types of education in different industries, or not?
3. Should a distinction be made between the same occupational classes in different industries, or not?

Which decisions are made, is closely related to the validity of both hypotheses. If the competitiveness hypothesis can not be rejected, types of education should be treated to be not industry-specific. If the technology hypothesis also can not be rejected neither occupational classes should be treated to be industry-specific. Furthermore, it may be reasonable to suggest that both hypotheses hold for clusters of industries rather than for the whole economy. The larger these clusters are, the higher the level of industrial aggregation level can be.

Of course the line of argument can be reversed. The performance (the quality of the forecasts) of RAS variants which are distinctive with regard to the aspects mentioned above gives some information about the validity of both hypotheses. This idea is used in this paper.¹⁴ Several variants of RAS are chosen in order to cover the links between the two hypotheses and the three aspects above.

Five industrial aggregation levels are distinguished. In the data used in this paper a distinction is made between 14 industries. The lowest aggregation level is therefore characterised by the fact that all these 14 industries are treated separately. On the higher aggregation levels the economy is treated to consist of respectively 7, 5, 3 and 1 (that is, the whole economy) clusters of industries.¹⁵

Given the degree of industrial aggregation, characterised by the number of industrial clusters o , a distinction is made between two main variants. Given the fact that data are used in which 49 occupational classes and 54 types of educations are distinguished, these variants can be described as follows.

14. Evans and Wabe have chosen a different approach. They transform (their variants of) the equations (16) and (17) into testable regression equations, in which they compare the RAS prediction of wage and production changes with the actual wage changes. Two further differences between their approach and the approach in this paper are: (1) they take wages literally, while in this paper the wage is treated to be a measure of the labour market scarcity and (2) they test RAS, while in this paper the focus is on the optimal aggregation level within RAS.

15. For more information about the industrial aggregation levels, see Appendix 1.

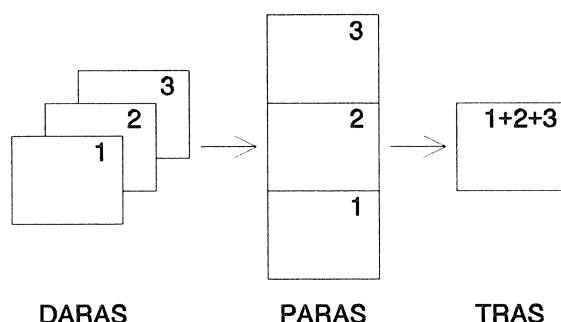
The RAS equivalent of the standard allocation model treats every industry separately. This implies that RAS is applied to 14 matrices of 49 rows by 54 columns. This variant is called *IRAS* (Industrial RAS). Besides this standard model, two variants of RAS are distinguished.

In the first variant RAS is applied to o separate 49 by 54 matrices. This variant implies that within the industries (or industrial clusters) both occupational classes and types of education are treated to be industry-specific. This variant is called *DARASo* (Disaggregated RAS with o industrial clusters, for example *DARAS3*). Two special cases of *DARAS* are distinguished: (1) *DARAS14* is exactly the same as *IRAS* and (2) *DARAS1*, which is called *TRAS* (Total economy RAS). In fact, *TRAS* is characterised by the fact that there is only one industrial cluster, both occupational classes and types of education are treated to be not industry-specific.

In the second variant RAS is applied to one ($o \times 49$) by 54 matrix. This variant implies that within the industries (or industrial clusters) only occupational classes are treated to be industry-specific. This variant is called *PARASo* (Partially Aggregated RAS with o industrial clusters). One special case of *PARAS* is distinguished; *PARAS1* which is exactly the same as *TRAS*.

To summarise, these variants can be shown in a figure:

Figure 1. RAS variants with $o = 3$



The *DARAS* variants are characterised by the fact that all (clusters of) industries are treated separately. In the *PARAS* variants the industrial matrices are linked in such a way that wage changes (that is, column totals) are not industry-specific. Finally, *TRAS* is characterised by the fact that the whole economy is treated to be one cluster.

Combining the industrial aggregation levels and the two main variants yields 9 RAS variants: *IRAS*, *DARAS7*, *DARAS5*, *DARAS3*, *PARAS14*, *PARAS7*, *PARAS5*, *PARAS3* and *TRAS*. However, in the light of investigating the hypotheses, the most important variants are *IRAS*, *PARAS14* and *TRAS*. The performance of these variants depends on two factors.

In the first place, the validity of the hypotheses. This is the theoretical factor. If both hypotheses can not be rejected, *IRAS*, *PARAS14* and *TRAS* will have the same performance. If only the competitiveness hypothesis holds, *TRAS* will perform worse than *PARAS14* and *IRAS*.

If both hypotheses can be rejected, *IRAS* will yield the best forecasts and *TRAS* will yield the worst forecasts.

In the second place, the noise in the data due to measurement errors. Because the data consist of samples, there is some noise. It is assumed that, the more the data are disaggregated, the greater the relative noise. So *IRAS* has to cope with the relatively greatest amount of noise and *TRAS* has the slightest trouble with noise. So:

let:

x_0, y_0, z_0 = theoretical performance *IRAS*, *PARAS14* and *TRAS*;¹⁶

x_1, y_1, z_1 = bias in the performance as a result of noise;

x_T, y_T, z_T = $x_0 + x_1, y_0 + y_1, z_0 + z_1$.

Suppose both hypotheses can not be rejected. Then:

Theory: $x_0 = y_0 = z_0$;

Noise: $x_1 > y_1 > z_1$;

Total: $x_T > y_T > z_T$.

Suppose only the competitiveness hypothesis holds. Then:

Theory: $x_0 = y_0 < z_0$;

Noise: $x_1 > y_1 > z_1$;

Total: $x_T > y_T \wedge y_T \geq z_T$.

Suppose both hypotheses can be rejected. Then:

Theory: $x_0 < y_0 < z_0$.

Noise: $x_1 > y_1 > z_1$.

Total: $x_T \geq y_T \wedge y_T \geq z_T$.

So, two conclusions can be drawn: (1) if *PARAS14* performs significantly better than *TRAS* (that is, $y_T < z_T$), the technology hypothesis can be rejected and (2) if *IRAS* performs significantly better than *PARAS14* (that is $x_T < y_T$), the competitiveness hypothesis can also be rejected. Furthermore, the optimal industrial aggregation level indicates the way in which industries are clustered with regard to the labour market and production technology.

Finally, it should be noted that these variants are executed for three sets of data: (1) forecasting 1981 based on the 1979 matrix, (2) forecasting 1983 based on the 1981 matrix

16. The better the performance, the lower the score x, y or z . See for the way in which the performance is measured section 7.

and (3) forecasting 1985 based on the 1983 matrix. The data set is composed of the Dutch labour force censuses (AKT), using numbers of people employed.

However, before focusing on the empirical analysis, some problems associated with RAS forecasting have to be solved. These problems occur in particular if highly disaggregated data are used to obtain information about the vectors r and s . These problems are discussed in the next section.

6. SOME PROBLEMS WITH RAS FORECASTING

6.1. The multiple solution problem

RAS does not yield a unique solution for the multiplying vectors r and s . If the vector r is multiplied and the vector s is divided by an arbitrary scalar, the same cell-forecasts c_{ij} would be obtained. For example, in section 2 it is mentioned that the RAS procedure can be initiated in two ways: by adjusting for the new row totals or by adjusting for the new column totals. These ways yield different multiplying vectors r and s . This problem is not relevant if RAS is used to obtain forecasts of the cells of the matrix, because RAS does yield unique cell forecasts. In this paper, however, the focus is on the multiplying vectors, which represent the wage changes. In terms of the allocation model this multiple solution implies that only relative price and wage changes can be derived. It is therefore necessary to normalise the wage vector in a useful way.

In this paper wages are normalised in the following way. Before starting the RAS procedure, the matrix A is multiplied by the change in the occupational employment (that is, the change in the row totals):

$$c_{ij}^{fc} = \frac{b_i^{occ}}{a_i^{occ}} \cdot a_{ij} \quad (30)$$

The outcome of this adjustment is the fixed coefficient forecast C^{fc} . This matrix can be interpreted as a forecast based on the assumption that, before demand-supply interactions take place, the educational structure of each occupational class does not change. This implies that, if the fixed coefficient method is applied, it is assumed that the scale of production does not have an impact on the educational structure. In the remainder of this paper RAS consists of two stages: (1) computing the fixed coefficient matrix C^{fc} and (2) executing the RAS procedure using C^{fc} as the base year matrix. This implies that the multiplying vectors r and s are determined by the second stage.

The way of normalising used here has several attractive features. Firstly, the forecasting procedure is not changed.¹⁷ In the second place, this procedure yields approximately stable multiplying vectors r and s . Finally, and most important here, Borghans and Heijke (1993) show that this procedure leads to a wage vector s for which approximately $\sum_j a_j^{educ} \cdot s_j = 0$, that is, wages are normalised such that on average the wage changes are 0.

6.2. The empty cells problem

Given a matrix for the base year and given row and column totals for the projection year, the RAS procedure does not always have a solution. This problem might occur if the base year

17. Note that, in fact, this adjustment is exactly the same as the first step of the RAS procedure which has been described in section 2.

matrix contains empty cells. The lower the aggregation level, the greater the probability that empty cells occur. Therefore, it is very important to investigate this problem in the context of this paper. Situations in which RAS does not have a solution are laid down in the following proposition:

Proposition 1

Let:

$$M = \{1..m\} \quad S_1 \in 2^M$$

$$N = \{1..n\} \quad S_2 \in 2^N$$

$$S_3^{S_2} = \{i \mid a_{ij} = 0 \quad \forall j \in S_2\}$$

$$S_4^{S_1} = \{j \mid a_{ij} = 0 \quad \forall i \in S_1\}$$

RAS does not have a solution if:

$$\exists S_1 \in 2^M \text{ such that } \sum_{i \in S_1} b_i^{occ} > \sum_{j \in S_4^{S_1}} b_j^{educ} \quad \vee$$

$$\exists S_2 \in 2^N \text{ such that } \sum_{j \in S_2} b_j^{educ} > \sum_{i \in S_3^{S_2}} b_i^{occ}$$

In words, the first part states the following: take a set of rows (the corresponding row numbers are in the set S_1 , which is a subset of M ; 2^M denotes the set of subsets of M) characterised by the fact that for certain columns in all these rows in the base year matrix A occur only empty cells. Subsequently, determine the set of columns in which all the cells belonging to these rows are empty (the corresponding column totals are in $S_4^{S_1}$). Suppose, in the target year matrix B , the sum of the row totals defined in S_1 ($\sum_{i \in S_1} b_i^{occ}$) is greater than the sum of the column totals in

which the cells in the rows concerned were not empty ($\sum_{j \in S_4^{S_1}} b_j^{educ}$). If such a set can be found,

RAS does not have a solution.¹⁸ The second part of the proposition can be described in a similar way. An example may clarify the problem:

Let:

$$A = \begin{pmatrix} 0 & 0 & 0 & 10 \\ 0 & 0 & 0 & 10 \\ 10 & 10 & 10 & 10 \\ 10 & 10 & 10 & 10 \end{pmatrix}$$

18. For a proof of proposition 1, see Appendix 2.

and:

$$B = \begin{pmatrix} 0 & 0 & 10 & 10 \\ 0 & 0 & 10 & 10 \\ 10 & 10 & 10 & 1 \\ 10 & 10 & 10 & 1 \end{pmatrix}$$

Take $S_1 = \{1,2\}$. As a result $S_4^{S_1} = \{1,2,3\}$. The rule in proposition 1 can be applied now:

$$\sum_{i \in S_1} b_i^{occ} = b_1^{occ} + b_2^{occ} = 20 + 20 = 40$$

$$\sum_{j \in S_4^{S_1}} b_j^{educ} = b_4^{educ} = 22$$

So if RAS is applied here and A is the base year matrix and B is the future matrix, no solution will be found. The reason is that in the forecasted matrix the empty cells of A will stay empty (that is, $c_{11} = c_{12} = c_{13} = c_{21} = c_{22} = c_{23} = 0$). This implies, for example, that both c_{14} and c_{24} have to be equal to 20 to satisfy the first and second row constraint. But, in that case, it is impossible to satisfy the fourth column constraint ($c_{14} + c_{24} = 40 > 22$).

Proposition 2 traces the situations in which the empty cells problem can occur, if the contents of the matrix B is known:¹⁹

Proposition 2:

The empty cell problem can occur if and only if:

$$\exists (i,j) \ i \in S_1, j \in S_4^{S_1} \text{ such that } b_{ij} > 0 \quad \forall$$

$$\exists (i,j) \ j \in S_2, i \in S_3^{S_2} \text{ such that } b_{ij} > 0$$

In words, only if there is an empty cell in matrix A that is not empty in matrix B , the empty cells problem can occur.

In order to solve the empty cells problem, empty cells in the matrix A have to be filled. Trying to find a solution for the empty cells problem leads to two questions. In the first place, which cells have to be filled? To find the right answer, given by proposition 2, in a 49 by 54 matrix, $2^{49} + 2^{54} = 18 \cdot 10^{15}$ sets have to be investigated in order to correct only for those situations in which the empty cells problem really might occur. This is an almost impossible task. In the second place, how should the empty cells be filled? If cells are filled with very small numbers, which seems to be attractive in order to remain the old structure as much as possible, the theoretical problem is replaced by a practical one. If very small numbers occur in the base year

19. For a proof of proposition 2, see Appendix 3.

matrix, the RAS solution converges very slowly and leads to very unsatisfying results and numerical problems.

Therefore, in this paper an 'optimal' correction number is searched for to fill every empty cell in the matrix *A*. Two criteria are used. Firstly, the number has to be relatively small compared to the volume of other cells in order to sustain the existing proportions as much as possible. Secondly, the number has to be relatively stable with regard to the performance of RAS. Because the minimum volume of a non-empty cell is 20,²⁰ the first criterium leads to the constraint that the number has to be smaller than 20. In order to take a final decision, the performance of *TRAS*, *PARAS14* and *IRAS* has been calculated for forecasts of 1985 based on data of 1983. The results are presented in the following table:²¹

Table 1. Scores with different correction numbers: 1983 - 1985

correction number	IRAS	PARAS14	TRAS
0.001	109,959.73	0.78	0.76
0.01	2,172.44	0.78	0.76
0.1	71.33	0.77	0.75
1	4.43	0.75	0.75
10	3.02	0.95	0.71
20	3.56	1.29	0.69
30	4.16	1.64	0.68

Table 1 shows that for numbers greater than 0.1, the performance does not improve substantially anymore or even gets worse (the higher the score, the worse the performance). Therefore the correction number is chosen to be 1. That is, every empty cell in the base year matrix is replaced by the number 1.

To summarise this section, the multiple solution problem is solved by calculating a fixed coefficient matrix before starting the RAS procedure. The empty cells problem is solved by replacing all empty cells in the base year matrix by 1. Given these solutions, in section 7 the empirical results will be analysed.

20. The reason is that the original data are samples. These data are multiplied afterwards to bring them in alignment with the volume of the total labour force. The maximum sample size is 5%, which implies that the contents of a cell is at least 20.

21. In table 1, the performance is reflected by the *score* of the RAS variant. For more information about the way in which the performance and the score are calculated, see section 7.1.

7. EMPIRICAL ANALYSIS

7.1. Measurement of the performance

A central item in this paper is the performance of the several RAS variants. The following criterium $Q(V)$, which is the sum of the relative quadratic forecast errors, is used to measure the performance:

$$Q(V) = \sum_{i,j: \max(a_{ij}, b_{ij}) > 0} \frac{(c_{ij} - b_{ij})^2}{\max(a_{ij}, b_{ij})} \quad \forall i, j \quad (31)$$

in which V is the variant involved. In order to avoid denominators equal to 0, the quadratic forecast errors are divided by the maximum of a_{ij} and b_{ij} instead of by a_{ij} or b_{ij} . This criterium has often been used to evaluate RAS forecasts.²² This performance measure $Q(V)$ is used to calculate the *score* of the variant V :

$$\text{score}(V) = \frac{Q(V)}{Q(SAB)} \quad (32)$$

in which $Q(SAB)$ is the performance of the so-called Same As Before method. This method consists of forecasting the matrix B by simply taking the matrix A as forecast. In other words, taking $c_{ij} = a_{ij}$. This specification implies that the greater *score* (V), the worse the performance of that variant V . $Q(SAB)$ can be considered to be lower constraint to the performance of the RAS variants. So, if *score*(V) appears to be significantly greater than one, the RAS variant V is not an improvement compared to SAB , which implies that it does not make much sense to apply this variant for manpower forecasting purposes.

Comparing the scores of different variants is the following step. Assuming that the relative quadratic forecast errors in equation (31) are normally distributed, it can be shown that:

$$F_{V_1, V_2} = \frac{\frac{\text{score}(V_1)}{df_1}}{\frac{\text{score}(V_2)}{df_2}}$$

in which V_1 and V_2 are two RAS variants and df_1 and df_2 are the degrees of freedom, has an F distribution with df_1 , df_2 degrees of freedom, if the variances of the quadratic forecast errors are equal for both forecast methods. Furthermore, $df_1 = df_2 = nec - m - n + 1$, in which nec is the number of non-empty cells, m is the number of rows and n is the number of columns.

22. See, for example, Evans and Lindley (1973) and Borghans and Heijke (1993). Furthermore, Borghans and Heijke (1993) give a theoretical and technical foundation for this criterium.

7.2. Analysis of the performances

Starting-point for the analysis is the performance of *SAB*, which is presented in table 2:

Table 2. Performance of *SAB*

years	Q(<i>SAB</i>)	Q(<i>SAB</i>) / <i>nec</i>
1979 - 1981	390,059.72	280.42
1981 - 1983	262,691.13	170.47
1983 - 1985	242,494.14	159.01

in which *nec* is the number of non-empty cells in the base year matrix before the adjustment. Based on the performance of *SAB*, the scores of the different RAS variants can be calculated (see equation (32)). The results are presented in table 3:

Table 3. Scores of the RAS variants

variant	'79-'81	'81-'83	'83-'85
IRAS	4.21	2.37	4.43
DARAS7	2.52	2.29	4.51
DARAS5	2.41	1.94	6.17
DARAS3	2.12	2.16	2.41
PARAS14	0.78	0.61	0.75
PARAS7	0.71	0.61	0.74
PARAS5	0.72	0.61	0.74
PARAS3	0.63	0.62	0.74
TRAS	0.69	0.63	0.75

As shown in section 5, comparing the score of *IRAS* and *PARAS14* offers the opportunity to test the competitiveness hypothesis. Furthermore, if the competitiveness hypothesis can not be rejected, comparing the score of *PARAS14* and *TRAS* offers the opportunity to test the technology hypothesis. Therefore, in table 4 the F-statistics are presented:

Table 4. Comparing *IRAS*, *PARAS14* and *TRAS*: F-statistics

V_1	V_2	'79-'81	'81-'83	'83-'85
IRAS	PARAS14	6.10*	3.76*	5.91*
PARAS14	TRAS	1.13*	0.97	1.00

* = significant (two-sided test, $\alpha = 0.05$)

Table 4 shows that $F_{IRAS, PARAS14}$ is significantly greater than one in all cases. In other words, *IRAS* performs significantly worse than *PARAS14*. This implies that the competitiveness hypothesis can not be rejected; wage changes are uniform over industries. Furthermore, table 4 shows that in all three cases *PARAS14* does not perform significantly better than *TRAS*. In one case (1979 - 1981) *PARAS14* performs even significantly worse. This implies that the technology hypothesis can not be rejected also, although the evidence is weaker; the occupational production technology is not industry-specific.

The analysis of the performance of the subvariants may give some information about clusters of industries with regard to the hypotheses. Table 3 shows that, with regard to the *PARAS* variants, the aggregation level does not have a significant impact on the performance. There is one exception: *PARAS3* in 1979 - 1981 performs significantly better than the other *PARAS* variants. With regard to the *DARAS* variants, table 3 shows a various picture. However, the highest aggregation level (that is, *DARAS1* = *TRAS*), appears to perform better than the other levels. Furthermore, all *DARAS* variants, except *TRAS*, perform significantly worse than *SAB*. So, it can be concluded that a partial clustering of industries rarely yields better forecasts than applying RAS to the economy as a whole. This is not a surprising result, because it is shown above that both the competitiveness hypothesis and the technology hypothesis can not be rejected for the Dutch economy as a whole.

7.3. The competitiveness hypothesis: further indications

It is shown above that the competitiveness hypothesis can not be rejected, because *PARAS14* performs significantly better than *IRAS*. In this section some other indications which may underpin the competitiveness hypothesis are presented.

Table 5. $\hat{\sigma}_{sj}^2$ and GH_j by type of education

Type of education	$\hat{\sigma}_{sj}^2$	GH_j
1 Primary Education	0.002	0.97
2 Lower General Secondary Education	0.004	0.94
3 Lower Vocational Education, Agriculture	0.027*	0.60
4 Lower Vocational Education, Technical	0.004	0.94
5 Lower Vocational Education, Transport & Harbour	0.036*	0.68
6 Lower Vocational Education, Commerce & Administration	0.012	0.90
7 Lower Vocational Education, Community Care, Hotel & Catering	0.007	0.89
8 Lower Vocational Education, Security	0.027*	0.84
9 Lower Vocational Education, n.e.c.	0.069*	0.89
10 Higher General Secondary Education	0.006	0.91
11 Intermediate Vocational Education, Agriculture	0.035*	0.57
12 Intermediate Vocational Education, Non-Medical laboratory	0.013	0.89
13 Intermediate Vocational Education, Engineering	0.004	0.94
14 Intermediate Vocational Education, Transport & Harbour	0.026*	0.69
15 Intermediate Vocational Education, Medical Laboratory	0.020	0.70
16 Intermediate Vocational Education, Nursing & Para-medical services	0.062*	0.22
17 Intermediate Vocational Education, Commerce & Administration	0.010	0.89
18 Intermediate Vocational Education, Administrative, legal & Fiscal	0.049*	0.35
19 Intermediate Vocational Education, Social & Cultural	0.024*	0.72
20 Intermediate Vocational Education, Community Care	0.024*	0.66
21 Intermediate Vocational Education, Hotel, Catering & Hairdressing	0.034*	0.45
22 Intermediate Vocational Education, Police, Fire & Defense Forces	0.061*	0.36
23 Intermediate Vocational Education, n.e.c.	0.016	0.74
24 Higher Vocational Education, Teacher Training	0.066*	0.35
25 Higher Vocational Education, Interpreter & Translator	0.057*	0.80
26 Higher Vocational Education, Theology	0.037*	0.39
27 Higher Vocational Education, Agriculture	0.013	0.88
28 Higher Vocational Education, non-medical Laboratory	0.010	0.92
29 Higher Vocational Education, Engineering	0.006	0.93
30 Higher Transport & Harbour	0.015	0.78
31 Higher Vocational Education, Medical Laboratory	0.091*	0.33
32 Higher Vocational Education, Nursing & Physiotherapy etc.	0.095*	0.18
33 Higher Vocational Education, Commerce & Administration	0.010	0.91
34 Higher Vocational Education, Business Administration Technology	0.017	0.88
35 Higher Vocational Education, Administrative, Legal & Fiscal	0.045*	0.63
36 Higher Vocational Education, Social & Cultural	0.043*	0.65
37 Higher Vocational Education, Hotel & Catering Industry	0.081*	0.65
38 Higher Vocational Education, Fine Arts	0.026*	0.74
39 Higher Vocational Education, Police, Fire & Defense Forces	0.041*	0.20
40 Higher Vocational Education, n.e.c.	0.014	0.91
41 Academic Education, Teacher Training	0.066*	0.28
42 Academic Education, Arts	0.051*	0.40
43 Academic Education, Theology	0.158*	0.40
44 Academic Education, Agriculture	0.008	0.83
45 Academic Education, Mathematics & Natural Sciences	0.024*	0.68
46 Academic Education, Engineering	0.008	0.89
47 Academic Education, Veterinary & Medical Sciences & Dentistry	0.068*	0.27
48 Academic Education, Pharmacy	0.015	0.75
49 Academic Education, Economics, Econometrics & Business Administration	0.013	0.81
50 Academic Education, Law & Public Administration	0.033*	0.69
51 Academic Education, Social Sciences	0.045*	0.64
52 Academic Education, Fine Arts	0.015	0.71
53 Academic Education, n.e.c.	0.226*	0.68
54 Education, n.e.c.	0.001	0.96

* = significant (one-sided test, $\alpha = 0.005$)

In table 5 the variance per type of education j of the weighted multiplying factors s_{jk} over industries k is shown for *IRAS* (1983 - 1985).²³ The variance is an indication for the competitiveness of the educational submarket j . The higher the variance, the more the wage changes vary over the industries and the less competitive the educational submarket is. Suppose the weighted multiplying factors are normally distributed. Then:

$$X_{n-1}^2 = \frac{(n-1) \hat{\sigma}_{sj}^2}{\sigma_0^2}$$

in which $\hat{\sigma}_{sj}^2$ is the sample variance, has a chi-square distribution with $(n-1)$ degrees of freedom. In table 5 a distinction is made between types of education characterised by a high variance (that is, significantly higher than 0.01)²⁴ on the one hand and types of education characterised by a low variance on the other. The types of education with a significantly high variance are marked with an asterisk.

Table 5 shows that 24 of the 54 types of education have a small variance. Among these are primary, lower and higher general secondary education, the technical, (including engineering and laboratory), the economic and some agricultural types. 19 of the 24 types are in these classes or are rest-categories (n.e.c.-classes). In 1985, the people having these 24 types of education ($= 24/54 = 44.4\%$) represent 75.1% of the total labour force.

In table 5 also a dispersion measure GH_j is presented. This measure shows how workers that finished type of education j are 'distributed' over the industries. GH_j , which is a variant of the Gini-Hirschman-index, is defined as follows (see De Grip and Heijke, 1988):

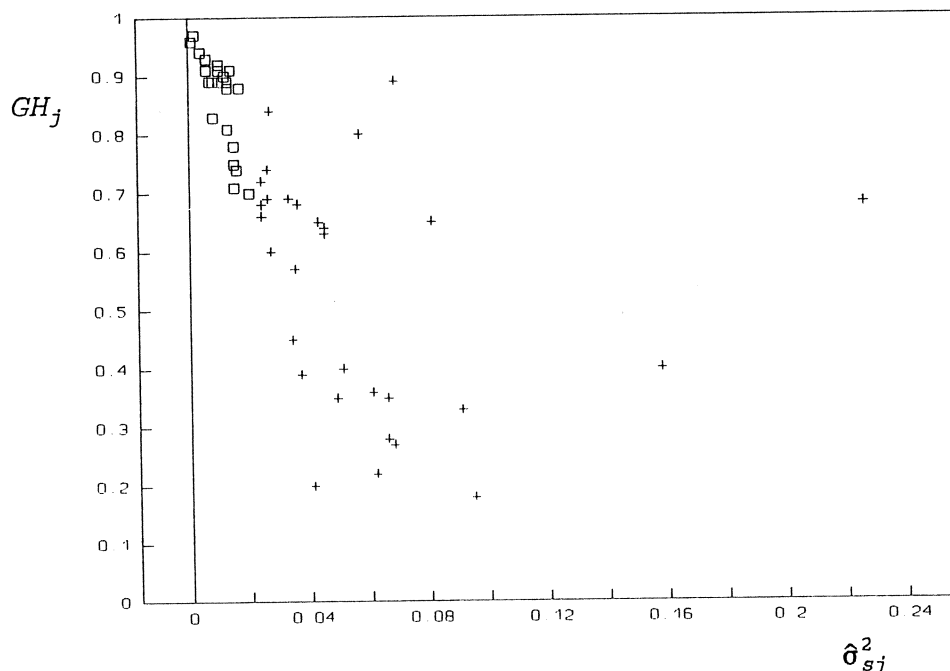
$$GH_j = \left(1 - \sum_k \left(\frac{L_{jk}}{L_j} \right)^2 \right) \left(\frac{o}{o-1} \right) \quad (33)$$

in which o is the number of industries. This measure is chosen, because its range is independent of o . GH_j is always between 0 and 1. If $GH_j = 0$, people having type of education j are concentrated in one industry. If $GH_j = 1$, people having type of education j are uniformly distributed over all industries. In figure 2 $\hat{\sigma}_{sj}^2$ and GH_j are presented in a scatter plot. In this figure the 24 types of education characterised by a low variance $\hat{\sigma}_{sj}^2$ are reflected by a \square , the types of education characterised by a high variance are reflected by a $+$. Figure 2 clearly

23. The weight factors give the importance of the sector for employment in that type of education. So the weight factor for s_{jk} is $\frac{L_{jk}}{L_j}$.

24. The null hypothesis $H_0: \sigma^2 \leq 0.01$ is arbitrarily chosen. However, the variance of s_{jk} over j and k is 0.035. This implies that types of education for which H_0 can not be rejected have a relatively small variance compared to the overall variance.

Figure 2. $\hat{\sigma}_{sj}^2$ and GH_j for types of education with low and high variance



\square = types of education with a low variance
 $+$ = types of education with a high variance

shows that both educational groups can be distinguished also with regard to the dispersion measure GH_j . All types of education with a low variance are clustered at the top of the figure. This implies that people having a type of education characterised by the fact that wage changes are relatively uniform over industries are relatively uniformly 'distributed' over industries. A Mann-Whitney test underpins this graphical finding. This test shows that the 24 types of education having a low $\hat{\sigma}_{sj}^2$ have a significantly higher mean GH_j than the types of education having a $\hat{\sigma}_{sj}^2$ which is significantly higher than 0.01.²⁵ This implies that people having these 24 types of education are significantly more uniformly distributed over the industries.

It can be concluded that about 75% of the labour force has a type of education which shows a relatively great uniformity in wage changes. Furthermore, these workers are relatively uniformly distributed over all industries. These findings can be considered to be indications to underpin the competitiveness hypothesis for (a large segment of) the labour market.

25. $n_1 = 24$, $n_2 = 30$, $Z = -6.14$, which is significant for a very small α ($\alpha < 0.0001$).

7.4. The aggregation problem

In RAS forecasting practice, the analysis in this paper can be a guideline for decisions on the optimal aggregation level. It has been shown that the performance of RAS depends on two factors: (1) the degree of uniformity of the multiplying vectors r and s over the industries and (2) the noise in the data. The higher the uniformity and the greater the amount of noise, the higher the optimal aggregation level. With regard to the Dutch data used in this paper, it can be concluded that the uniformity of the vectors r and s is relatively high and the amount of noise is relatively great. This results in a high optimal aggregation level. It should be noted here that, in the analysis of this paper, the row and column totals of the future matrix were known. If these totals have to be forecasted (as in normal practice) even more noise may occur, especially on lower aggregation levels.

Of course it is difficult to draw general conclusions with regard to the aggregation problem. However, some remarks can be made. Complete uniformity in the multiplying vectors r and s will never occur. This implies that, from a theoretical point of view, disaggregation always leads to better forecasts. So the real problem is the quality of the data. In general, the lower the uniformity of industries and the better the quality of the data are, the more scope there is for disaggregation. Some examples can be given. Firstly, if the sample used or the labour market which is investigated (for example, a country, region or industry) is small, the amount of noise will be relatively great. This implies that it may be recommendable then to choose a high aggregation level. In the second place, evidence on the issues incorporated in the competitiveness hypothesis and the technology hypothesis may be a guideline for decisions on the aggregation level and the RAS variant which is used. For example, if some empirical evidence is available on the uniformity of developments in the degree of market scarcity at the labour market which is investigated, a *PARAS* variant may be recommendable. Finally, the quality of the future occupational and educational structure (that is, the future row and column totals) has to be considered. If these forecasts are very poor and assuming that the aggregated structure is easier to predict than disaggregated structures, it may be recommendable to choose a high aggregation level. The empirical analysis in this paper based on Dutch data shows that the uniformity of wage over industries is a reasonable assumption. The evidence on the uniformity of production functions, on the other hand, is less convincing, however.

8. CONCLUSIONS

In this paper an allocation model is used to give an economic foundation of the RAS model in a manpower forecasting context. Starting-point for the allocation model is a two-stage production function. In the first stage the way in which capital and occupational activities generate output is described. In the second stage it is shown how workers having particular qualifications generate the output by occupational activity. Solving the profit maximising problem leads to a labour demand equation which can be re-arranged to obtain a RAS model.

From a theoretical point of view, the allocation model offers the opportunity to investigate the validity of two hypotheses concerning the labour market. The so-called competitiveness hypothesis states that wage changes for each type of education are uniform over the market. This can be considered to be an indication for the competitiveness of the labour market. The so-called technology hypothesis states that the production function of occupational activities is not industry-specific. This implies that the contents of occupational activities are uniform over all industries. That is, the production function which links the use of different types of education to the occupational activities is equal in every industry. Furthermore, it implies that the productivity of workers only depends on their qualifications and the occupational activity they are employed in. From a practical point of view a guideline can be given for decisions on the optimal aggregation level. If both hypotheses can not be rejected for a large segment of the economy, a high aggregation level may be recommendable.

It is shown that variants of the allocation model with regard to both hypotheses can be represented by variants of RAS. Besides the standard model (*IRAS*), in which both hypotheses are dropped, which implies that both row and column totals are disaggregated by industry or clusters of industries, two main variants of RAS are distinguished. In the first variant (*PARAS14*) the competitiveness hypothesis is accepted. This implies that, on the one hand industrial occupational classes are distinguished, while on the other hand types of educations are aggregated over industries. In other words: row totals are disaggregated by industry, while column totals are not. In the second variant (*TRAS*) both hypotheses are accepted. This implies that both row and column totals are aggregated over industries. Furthermore, these two variants are applied to several aggregation levels.

From a microeconomic point of view, the empirical analysis based on Dutch data shows that both the competitiveness and the technology hypothesis can not be rejected for the Dutch economy. Because *PARAS14* gives significantly better forecasts than *IRAS*, the competitiveness hypothesis can not be rejected. Because *TRAS* does not give significantly worse forecasts than *PARAS14*, the technology hypothesis can not be rejected. The performance of the RAS variants with regard to the distinguished aggregation levels shows that the highest level does not give significantly worse forecasts than the other levels. This implies that both hypotheses can not be rejected for the Dutch economy as a whole.

The majority of the labour force has a type of education which shows a relatively great uniformity in wage changes. Furthermore, these people are relatively uniformly distributed over all industries. These findings can also be considered to be indications to underpin the competitiveness hypothesis for (a large segment of) the labour market.

From a practical point of view, it should be noted that, in principal, disaggregation leads to better forecasts. So, the real problem is the quality of the data. In general, the lower the uniformity of the industries and the higher the quality of the data are, the more scope there is for disaggregation. For example, it may be concluded that (1) the smaller the sample or labour market considered, (2) the more evidence on the hypotheses is given and (3) the lower the quality of the forecasts of the future occupational and educational structure, the stronger a high aggregation level may be recommended. For the Dutch situation it is concluded that there is no empirical justification for incorporating wage-differentials between different industries for one type of education. In these data there is no need to distinguish differences in the production function between the industries also. This empirical observation is much less strong, however.

REFERENCES

Borghans, L. and Heijke, H. (1993), *Forecasting the Educational Structure of Occupations: A Manpower Requirement Approach with Substitution*, ROA-RM-1993/2E, Maastricht.

Bosworth, D.L., Evans, G.J. and Lindley, R.M. (1974), 'Mechanic Manpower Models', in: Wabe, J.S. (ed), *Problems in Manpower Forecasting*, Saxon House, Westmead, pp. 61-84.

Brown, A., Leicester, C. and Pyatt, G. (1964), 'Output, Manpower and Industrial Skills in the United Kingdom', in: *The Residual Factor and Economic Growth*, OECD, Paris.

Cohen, S.I (1988), 'Manpower Planning Models with Labour Market Adjustments: Applications to Columbia, Republic of Korea and Pakistan', *Economic Modelling*, vol. 5, january, pp. 19-31.

Corcoran, T., Hughes, G. and Sexton, J.J. (1993), *Occupational Employment Forecasts 1996*, report no. 3, FAS/ESRI, Dublin.

De Grip, A. and Heijke, H. (1988), *Labour Market Indicators: an Inventory*, ROA-W-1988/1E, Maastricht.

Evans, G.J. and Lindley, R.M. (1973), 'The Use of RAS and Related Models in Manpower Forecasting', *Economics of Planning*, vol. 13., no. 1-2, pp. 53-73.

Evans, G.J and Wabe, J.S. (1974), 'Testing a Demand Explanation of the RAS Model', in: Wabe, J.S. (ed), *Problems in Manpower Forecasting*, Saxon House, Westmead, pp. 85-95.

Van Eijs, P. (1993), *The Manpower Requirements Approach, Background and Methodology*, ROA-RM-1993/3E, Maastricht.

APPENDIX 1. INDUSTRIAL AGGREGATION LEVELS

Five aggregation levels are defined in which respectively 14, 7, 5, 3 and 1 sector are distinguished. In the standard model the lowest aggregation level is used. The following fourteen sectors are distinguished:

1. Agriculture, fishing, forestry.
2. Manufacture of foodstuffs, beverages, tobacco products.
3. Manufacture of textiles and footwear.
4. Other manufacture.
5. Chemical industry.
6. Metal industries.
7. Other industries.
8. Mining and quarrying.
9. Construction.
10. Trade.
11. Transport and communication.
12. Other private services.
13. Public services.
14. Public administration

In the second aggregation level seven sectors are distinguished (between brackets the original sectors which are joined):

1. Agriculture, fishing, forestry (1).
2. Manufacture (2,3,4).
3. Industry and mining (5,6,7,8).
4. Construction (9).
5. Trade (10).
6. Private services (11,12).
7. Public services and administration (13,14).

In the third level five sectors are distinguished:

1. Agriculture, fishing, forestry (1).
2. Manufacture and industry (2,3,4,5,6,7,8).
3. Construction (9).
4. Trade and private services (10,11,12).
5. Public services and administration (13,14).

In the fourth level three sectors are distinguished:

1. Agriculture, fishing, forestry (1).
2. Manufacture, industry and construction (2,3,4,5,6,7,8,9).
3. Trade, services and administration (10,11,12,13,14).

Finally, in the highest aggregation level the whole economy is considered as one sector.

APPENDIX 2. PROOF OF PROPOSITION 1

Suppose the matrix B has the property that:

$$\exists S_1 \in 2^M \text{ such that } \sum_{i \in S_1} b_i^{occ} > \sum_{j \in S_4^{S_1}} b_j^{educ} \quad (A2.1)$$

Furthermore, suppose that:

$$S_4^{S_1} \neq \emptyset \quad (A2.2)$$

If RAS has a solution, this solution would have the following characteristics:

$$c_i^{occ} = b_i^{occ} \quad \forall i \quad (A2.3)$$

$$c_j^{educ} = b_j^{educ} \quad \forall j \quad (A2.4)$$

$$a_{ij} = 0 \Rightarrow c_{ij} = 0 \quad \forall i, j \quad (A2.5)$$

From (A2.3), (A2.4) and (A2.5):

$$\sum_{i \in S_1} b_i^{occ} = \sum_{i \in S_1} \sum_j c_{ij} = \sum_{i \in S_1} \sum_{j \in S_4^{S_1}} c_{ij} \leq$$

$$\sum_i \sum_{j \in S_4^{S_1}} c_{ij} = \sum_{j \in S_4^{S_1}} c_j^{educ} = \sum_{j \in S_4^{S_1}} b_j^{educ}$$

So, if RAS a solution, then:

$$\sum_{i \in S_1} b_i^{occ} \leq \sum_{j \in S_4^{S_1}} b_j^{educ} \quad (A2.6)$$

This contradicts with assumption (A2.1). Therefore, RAS has no solution if the matrix B has property (A2.1).

If the matrix B has the property that:

$$\exists S_2 \in 2^N \text{ such that } \sum_{j \in S_2} b_j^{educ} > \sum_{i \in S_3^{S_2}} b_i^{occ}$$

it can be proven analogously that RAS has no solution.

APPENDIX 3. PROOF OF PROPOSITION 2

Consider a set:

$$S_1 \in 2^M \text{ such that } S_4^{S_1} \neq \emptyset$$

Then, the following equations can be derived:

$$\sum_{i \in S_1} b_i^{occ} = \sum_{i \in S_1} \sum_{j \in S_4^{S_1}} b_{ij} + \sum_{i \in S_1} \sum_{j \in S_4^{S_1}} b_{ij} \quad (A3.1)$$

$$\sum_{j \in S_4^{S_1}} b_j^{educ} = \sum_{i \in S_1} \sum_{j \in S_4^{S_1}} b_{ij} + \sum_{i \in S_1} \sum_{j \in S_4^{S_1}} b_{ij} \quad (A3.2)$$

From (A3.1) and (A3.2):

$$\begin{aligned} \sum_{i \in S_1} b_i^{occ} &> \sum_{j \in S_4^{S_1}} b_j^{educ} \rightarrow \\ \sum_{i \in S_1} \sum_{j \in S_4^{S_1}} b_{ij} &> \sum_{i \in S_1} \sum_{j \in S_4^{S_1}} b_{ij} - \\ \sum_{i \in S_1} \sum_{j \in S_4^{S_1}} b_{ij} &> 0 \rightarrow \\ \exists (i,j) \ i \in S_1, j \in S_4^{S_1} \text{ such that } b_{ij} > 0 \end{aligned} \quad (A3.3)$$

If a set:

$$S_2 \in 2^N \text{ such that } S_3^{S_2} \neq \emptyset$$

is considered, the proof is analogous.